

**GATHERING ENRICHED WEB SERVER
ACTIVITY DATA OF CACHED WEB CONTENT**

5

BACKGROUND OF THE INVENTION

The present invention relates generally to client-server computer systems and, more specifically, to information access requests to a web site server over a global communications network.

10

All web pages are written with HyperText Markup Language (HTML). Hypertext and universality are two essential features of HTML. Hypertext means that a programmer can create a link on a web page that leads the visitor to any other web page or to practically anything else on the Internet. Hypertext enables information on the web to be accessed from many different directions. Universality means that because HTML documents are saved as ASCII or text only files, virtually any computer can read a web page. HTML lets the web designer format text, add graphics, sound, and video, and save it all in a text or an American Standard Code for Information Interchange (ASCII) file that any computer can read. The key to HTML is in the tags, which are key words enclosed between less than (<) and greater than (>) signs, that indicate the type of content coming up next. While practically any computer can display web pages, how those pages actually look

15

depends on the type of computer, the monitor, the speed of the Internet connection, and the browser software used to view the page.

Advanced web designers often use a scripting language called JavaScript and a system of naming parts of the web page called the document object model (DOM), together with HTML to create dynamic content on a page. These effects are sometimes called dynamic HTML, or DHTML. HTML tags are commands written between angle brackets (< >) that indicate how the browser should display the text. Examples of HTML tags are BASE, FORM, FRAME, IMG and SCRIPT. There are opening and closing versions for many tags and the affected text is contained within the two tags. The opening and closing tags use the same command word; the closing tag carries an initial forward slash (/) symbol. Many tags have special attributes that offer a variety of options for the contained text. The attribute is entered between the command word and the final angle bracket. A series of attributes can be used in a single tag just by writing one after the other, in any order, with a space separating each one. The attributes in turn, often have values. In some cases, a selection of value is made from a small group of choices. Other attributes are more strict about the type of values they accept. Examples of attributes are HREF, SRC, ACCESSKEY and VALUE.

A web page is nothing more than a text document written with HTML tags. Like any other text document, web pages have a file name that identifies the documents to the web site designer, the web site visitors, and a visitor's web browser. Uniform Resource Locators (URLs) contain information about where a file is located and what a browser should do with it. Each file on the Internet has a unique URL. The first part of the URL is called the scheme. It tells the browser how to deal with the file that it is about to open. One of the most common schemes to access web pages is Hypertext Transfer Protocol (HTTP). The second part of the URL is the name of a server where

the file is located followed by the path that leads to the file and the file name. Sometimes, a URL ends in a trailing forward slash with no file name given. In this case, the URL refers to the default file in the last directory in the path (i.e., index.html), which generally corresponds to the home page. For example, consider the web address "census.rolandgarros.org/rc/images/...". The domain name is "census.rolandgarros.org". This is the specific host computer on which corresponding web pages reside. The next segment of the URL is the directory ("rc") and subdirectory "images") on the host computer that contains a specific web site. The last segment of the URL, represented by the ellipsis mark, is the filename of the specific web page being requested.

URLs can be either absolute or relative. An absolute URL shows the entire path to the file, including the scheme, server name, the complete path, and the file name itself. A relative URL describes the location of the desired file with reference to the location of the file that contains the URL itself. The relative URL for a file that is in the same directory as the current file is simply the file name and extension.

To view a single page, the browser running on a client computer, may request and download numerous files from a web site server. The number of object access requests ("hits") stored in the web site server's access log will typically exceed the number of distinct client sessions in which clients are accessing information on the web site, reducing the accuracy of the access log.

Data networking is growing at a phenomenal rate. The number of web users is expected to increase by a factor of five over the next few years. The resulting uncontrolled growth of web access requirements is straining all attempts to meet the bandwidth demand. Additionally, although the volume of web traffic on the Internet is staggering, a large percentage of that traffic is redundant, i.e., multiple users at any given site request much of the same content. This means that a significant

percentage of the wide area network (WAN) infrastructure carries the identical content and identical requests for accessing it daily. Web caching performs a local storage of web content to serve these redundant user requests more quickly, without sending the requests and the resulting content over the wide area network.

5 Caching is the technique of keeping frequently accessed information in a location close to the requestor. A web cache stores web pages and content on a storage device that is physically or logically closer to the user. This access to stored web content is closer and faster than a web lookup. By reducing the amount of traffic on wide area network links and on already overburdened web servers, caching provides significant benefits to Internet Service Providers (ISPs), enterprise networks, and end users. The two key benefits of web caching are cost savings due to the reduction of WAN bandwidth and improved productivity for end users resulting from quicker access. ISPs can place cache engines at strategic points on their networks to improve response times and lower the bandwidth demand on their backbones. ISPs can station cache engines at strategic WAN access points to serve web requests from local storage, rather than from a distant or overburdened web server. In enterprise networks, the dramatic reduction in bandwidth usage due to web caching allows a lower bandwidth WAN link to service the user base. Alternatively, the organization can add users or add more services that make use of the free bandwidth on the existing WAN link. For the end user, the response of the local web cache is almost three times faster than the download time for the same content over the wide area network. Therefore, users see dramatic improvements in response times, and the implementation of web caching is completely transparent to them.

Web caching offers other benefits including access control, monitoring and operational logging. The cache engine provides network administrators with a simple, secure method to enforce

a sitewide access policy through Uniform Resource Locator (URL) filtering. Network administrators can learn which URLs receive hits, the number of hits per second the cache is serving, the percentage of URLs that are served from the cache, along with other related operational statistics.

Web caching starts by an end user accessing a web page over the Internet. While the page is being transmitted to the end user, the caching system saves the page and all of its associated graphics on local storage. The page content is now cached. Another user, or the original user can then access the web page at a later time, but instead of sending the request over the Internet to the web server, the web cache system delivers the web page from local storage. This process speeds download times for the user, and reduces the bandwidth demand on the WAN link. Updating of the cache data can occur in a number of ways depending upon the design of the web cache system.

Web caching can be a major problem for publishers of web content. For example, a publisher can gather an inaccurate number of hits if some of the visitors access web content already in a caching server. Furthermore, if a caching server doesn't update content promptly, it can return expired or stale content to users.

SUMMARY OF THE INVENTION

Cache engines are becoming pervasive on the World Wide Web. As a result, the origin web servers do not serve or see the majority of the user requests for web site content. Packet sniffers will not see the requests either, as they are satisfied by cache engines elsewhere on the Internet. The technique of using a single pixel clear GIF (which is not cacheable) has been used to ensure that some record is recorded by the origin server for advertisements for some years. However, this solution only logs information about the request for the single pixel GIF file itself.

The single-pixel transparent GIF (Graphic Interchange Format) is the most flexible tool in a web designer's toolbox. The use of a transparent GIF is a way to discretely control the layout of text and graphics on the web page. No matter where the transparent GIF is placed on the page, it will remain unseen with all background graphics and fills remaining untouched. The single pixel clear GIF has been used before, but the data has not been enriched such that it can be used as a surrogate for the complete set of log records.

The present invention enriches the information recorded in the web logs for the uncacheable single pixel clear GIF by appending additional information to it as Common Gateway Interface (CGI) query string parameters. This enables the log record created by the request for the single pixel clear GIF to function as a "surrogate" for the complete set of log records which would have been created if the page content had not been cached.

DESCRIPTION OF THE DRAWINGS

The invention is better understood by reading the following detailed description of the invention in conjunction with the accompanying drawings, wherein:

Fig. 1 illustrates an implementation of web cache engines over a global communications network.

Fig. 2 illustrates an exemplary implementation of the uncacheable single pixel GIF with CGI query string parameters added to enrich information recorded in web logs.

Fig. 3 illustrates the processing logic for handling client requests for web pages utilizing the single-pixel transparent GIF in accordance with a preferred embodiment of the present invention.

Fig. 4 illustrates a site level analysis display that can be generated based on the implementation of the single pixel transparent GIF of the present invention.

Fig. 5 illustrates an exemplary display of referral categories that can be generated based on the implementation of the single pixel transparent GIF of the present invention.

5 Fig. 6 illustrates an exemplary display of referral category for search engines and directories that can be generated based on the implementation of the single pixel transparent GIF of the present invention.

10 Fig. 7 illustrates an exemplary display of the referral results for a specific search engine that can be generated based on the implementation of the single pixel transparent GIF of the present invention.

Fig. 8 illustrates exemplary content categories for various web pages that can be generated based on the implementation of the single pixel transparent GIF of the present invention.

Fig.9 illustrates an exemplary content category for a home page that can be generated based on the implementation of the single pixel transparent GIF of the present invention.

15 Fig. 10 illustrates an exemplary display of the available saved reports that can be generated based on the implementation of the single pixel transparent GIF of the present invention.

Figs. 11A - 11M illustrate various available saved reports that can be generated based on the implementation of the single pixel transparent GIF of the present invention.

20 **DETAILED DESCRIPTION OF THE INVENTION**

Web server software typically collects and saves information pertaining to each HTTP request, including date and time, the originating Internet Protocol (IP) address, the object requested,

and the completion status of the request. The logs are analyzed on a periodic basis to determine the traffic through the server in terms of hits, the number of pages served, and the level of demand for pages of interest during each period.

Internet browser applications allow an individual user to cache web pages on his local hard disk. A user can configure the amount of disk space devoted to caching. The first time a user views a website, that content is saved as files in a subdirectory on that computer's hard disk. The next time the user points to this website, the browser gets the content from the cache without accessing the network. Certain elements of the page, including buttons, icons and images, appear much more quickly than they did the first time the page was opened.

To limit bandwidth demand caused by the uncontrolled growth of Internet use, software developers have developed applications that extend local caching to the network level. The two current types of network level caching products are proxy servers and network caches. Proxy servers are software applications that run on general-purpose hardware and operating systems. A proxy server is placed on hardware that is physically between a web browser client application and a web server. The proxy server acts as a gatekeeper that receives all the packets destined for the web server and examines each packet to determine whether it can fulfill the request itself. If the proxy cannot fulfill the request itself, it forwards the request to the web server. Proxy servers can be used to filter requests, e.g., to prevent employees from accessing specific websites. The problem with using proxy servers is that they are not optimized for caching and can fail under a heavy network load. Traffic is slowed to allow the proxy servers to examine each packet, and the failure of the proxy software or hardware causes all users to lose network access. Furthermore, proxy servers require configuration of each end-user's browser, which is an unacceptable option for ISPs and large

enterprises. Because of these shortcomings of proxy servers, applications that create network caches have become popular. These caching-focused software applications are designed to improve performance by enhancing the caching software and eliminating the other slow aspects of proxy server implementations. Because a proxy server is run under a general purpose operating system that involves very high per-process context overhead, they are not easily scaleable to large numbers of simultaneous processes.

Networking product vendors offer cache engines as a single purpose network appliance that stores and retrieves content using caching and retrieval algorithms. Such cache engines are dedicated solely to content management and delivery. Since only web requests are routed to the cache engine, no other user traffic is affected by the caching process. For non-web traffic, the router functions entirely in its traditional role. The communications between a cache engine and a router is defined by a cache control protocol. Under this protocol, the router directs only web requests to the cache engine rather than to the intended server. With a cache engine, a client requests web content in the usual manner. A router running a cache control protocol intercepts Transmission Control Protocol (TCP) port 80 web traffic and routes it to the cache engine. The client is not involved in the transaction, and no changes to the client or browser are required. If the cache engine does not have the requested content, it sends the request to the Internet or Intranet in the usual fashion. The content is returned to and stored at the cache engine. The cache engine returns the content to the client. Upon subsequent requests for the same content, the cache engine fulfills the requests from local storage.

Fig. 1 illustrates an implementation of web cache engines over a global communications network such as the Internet. A client computer 12, 14, 16 can request web content via a router 18.

The router 18 intercepts TCP Port 80 web traffic and routes it to the local cache engine 20. The client 12, 14, 16 is not involved in this transaction and no changes to the client computer or browser are required. If the cache engine 20 does not have the requested content, it sends the request via router 18 to the Internet to access an Internet content server 40, 42, 44. The content is returned to, and stored at, the cache engine 20. The cache engine 20 then returns the requested content to the client computer 12, 14, 16 via the router 18. Several cache engines 32, 34, 36 can be placed in a cache farm in a hierarchical fashion at an Internet Service Provider (ISP) site 30. Requests from clients 12, 14, 16 directed through router 18 and ISP server 30, are diverted to the cache farm 32, 34, 36 to fulfill the client request from its storage. If the cache engines 32, 34, 36 are unable to fulfill the request from local storage, a normal web request is made via ISP server 30 over the Internet to an appropriate server 40, 42, 44 for the requested Internet content. In addition to router 18, routers 26, 46 are also shown connected to ISP server 30. Routers 18, 26, 46 are frequently referred to as Points-of-Presence (POPs). A POP is the location of an access point to the Internet and has a unique Internet IP address. A POP usually includes routers, digital/analog call aggregators, servers and frequently frame relay or Asynchronous Transfer Mode (ATM) switches. Shown connected to router 46 is cache engine 48. Connected to router 26 is cache engine 28 and router 24. Router 24 is connected to a corporate intranet 22.

Because the router redirects packets destined for web servers to the cache engine, the cache engine operates transparently to clients. Clients do not need to configure their browsers to be in proxy server mode. In addition, the operation of the cache engine is transparent to the network. The router operates entirely in its normal role for non-web traffic.

A web object can contain a Hypertext Transfer Protocol (HTTP) header to instruct a browser in a caching server how to cache the web object. For a static image, such as a company logo, the expiration header can be set to "no expiration" so that caching servers can keep the image in the cache forever. In order to gather the exact number of hits on a specific page, e.g., an advertisement, a small image object can be added to the page with the object set to expire immediately, so the caching server won't cache the object. Then, every time a user requests that page, the browser or caching server will retrieve the object from the original web server, and the web server can then count the exact number of requests.

The Common Gateway Interface (CGI) is a simple interface (protocol) for running external programs, software or gateways under an information server in a platform-independent manner. CGI is simply a standardized way for sending information between the server and the script. The CGI script is a program that communicates with the server in a standard way. Currently, the supported information servers are HTTP servers. Each CGI server implementation must define a mechanism to pass data about the request from the server to the script.

Each element on a web page form will have a name and value associated with it. The name identifies the data being sent. The value is the data and can either come from the web page designer or from the visitor who types it in a field. When a visitor clicks the submit button, the name - value pair of each form element is sent to the server. CGI scripts generally have two functions. The first is to take all the name-value pairs and separate them out into individual intelligible pieces. The second is to actually do something with that data, such as printing it out, multiplying fields together, sending an email confirmation, or storing it on a server. The form has three important parts: the form tag, which includes the URL of the CGI script that will process the form; the form elements,

such as fields and menus; and the submit button which sends the data to the CGI script on the server.

Scripts are little programs that add interactivity to a web page. Simple scripts can be written to add an alert box or some text to the web page; more complicated scripts can be written that load particular pages according to the visitor's browser or that change a frame's background color depending on the visitor's mouse clicks. Most scripts are written in a scripting language called JavaScript that is supported by most browsers, including Netscape Communicator and Microsoft Internet Explorer.

JavaScript is an object-oriented language, which means that it works by manipulating objects on a web page, such as windows, images and documents. JavaScript commands are put directly into the HTML file that creates a web page. Depending on the script being run, the commands can be placed into several parts of the file. The commands are frequently placed near the top of the file. Special codes set off the commands, alerting the browser that they are JavaScript commands. If the commands are put before the HTML <Body> tag at the top of the file, then the script will be able to start executing while the HTML page is still loading. JavaScript is an interpreted language, which means its commands are executed by the browser in the order in which the browser reads them. JavaScript works by taking actions on objects. These actions are called methods. In the basic syntax of JavaScript, the object is first named, and then a period appears followed by the action taken on the object, i.e., the method. So the command to open a new window in JavaScript is window.open. In this instance, window is the object and open is the method. This command opens a new browser window. Other parameters can be added after the command. All the parameters are placed inside one set of parenthesis, with each individual parameter inside quotation marks, with the parameters separated by commas.

An automatic script is executed by the client browser when the web page is loaded. There is no limit to the number of automatic scripts that can be on a web page. The location of the script on the HTML page determines when the script will load. Scripts are loaded in the order in which they appear in an HTML document. An automatic Java Script is added to an HTML document by the following HTML code:

```
<SCRIPT LANGUAGE="JavaScript">  
type content of the script  
</SCRIPT>
```

Some of the older browsers cannot run scripts and will not understand the SCRIPT tag. In order to provide information to a visitor accessing an HTML page, an alternate way to provide information is through the use of the NOSCRIPT tag, followed by the information that is treated as regular text. The older browser won't understand the NOSCRIPT tag and will ignore it, but process the following text. The following is added to the HTML document:

```
<NOSCRIPT>  
type the information  
</NOSCRIPT>
```

In the implementation of the single pixel GIF to create surrogate log files, the following tags and attributes are used as illustrated in Fig 2 discussed below:

IMG is the HTML tag for inserting images on a page;

ALT is an attribute for offering alternate text that is displayed if the image is not;

SRC is an attribute for specifying the URL of the image;

Also illustrated in Fig. 2 are the following attributes for the IMG tag:

WIDTH, HEIGHT are attributes for specifying the size of the image so that the HTML page can be loaded more quickly;

BORDER is an attribute for specifying the thickness of the border, if any. BORDER=0 omits the border that a browser would otherwise place automatically around an image.

5 In a preferred embodiment of the present invention, a CGI string of data is appended to the SRC attribute for the single pixel GIF at the time the page is published, as follows:

&pag=xxxxxxx the absolute URL of the page on which the GIF appears;
&num=xx the number of elements (SRCs) on the page at the time of publishing;
10 &ref=xxxxxxxxx the URL of the page which requested the current page (this is done via Java Script).

In addition, the persistent cookie identification of the user's cookie can be appended to the CGI string of data as follows:

&usr=xxxxxxx the persistent cookie ID of the user cookie (Java Script).

15 Fig. 2 illustrates an example of an implementation of the single pixel GIF with the addition of query string parameters to act as a surrogate for the complete set of log records that would have been created had the page content not been cached. In Fig. 2, the Java Script statements are embedded directly on the HTML page. It includes a document object with a write method ("document.write"). The document object contains information on the current document and provides methods for displaying the HTML expressions to the user in a specified window. The IMG
20 and BR tags are the HTML expressions that are displayed in the window. The BR CLEAR tag and attribute simply create a line break and stop text wrap. The SRC attribute following the IMG tag provides the absolute URL of the page containing the single pixel clear GIF ("uc.GIF"); i.e.,

SRC="http://census.rolandgarros.org/rc/images/uc.GIF?pag=' + location.pathname
+ '&num=14' + r + '".

The CGI string following uc.GIF indicates that there are 14 SRC elements on the HTML page. The URL of the referrer page is indicated by a variable "r", which is defined as '&ref='
5 +top.document.referrer based on a true condition to the "if" statement (i.e., the document referrer object is not empty). The Java Script top.document.referrer reflects the URL of the calling document (i.e., referrer page) that the user was viewing before the current page.

In the event the client browser cannot interpret a scripting language, the NOSCRIPT tag demarcates the HTML statements to be interpreted by the browser. This includes the IMG tag
10 wherein the SRC attribute has a query string after "uc.GIF" that is modified to include the default URL of the HTML page (i.e., "index.html"). The index.html file is the default file for the top level directory on the web site.

In order to serve up web pages, web sites need a host computer and server software that runs on the host. The host manages the communications, protocols, and houses the pages and related
15 software required to create a website on the Internet. The server software resides on the host and serves up the pages, and otherwise acts on the requests sent by the client's browser software. The server handles the HTTP requests and communications with the host operating system, which in turn handles the TCP/IP communications. There are different types of server software that perform different types of services for different types of clients. Specifically, a web server is an HTTP server
20 and its function is to send information to the client software (browser) using the HyperText Transfer Protocol. The client browser requests that the server return an HTML document. The server receives this request and sends back a response. The top portion of the response includes

transmission information and the rest of the response is the HTML file. In addition to sending pages to the browser, a web server also passes requests to run CGI scripts to CGI applications. These scripts run external mini-programs, such as a database lookup or interactive forms processing. The server sends the script to the application via CGI and communicates the script back to the browser.

5 The server software also includes configuration files and utilities to secure and manage the website in a variety of ways.

Fig. 3 illustrates the processing logic of the present invention. The process starts in start block 300. In logic block 302, the client browser software requests an HTML web page. The client browser determines if the requested HTML page has been cached at the client in decision block 302. If the page has been cached at the client, then the HTML file is delivered to the browser as indicated in logic block 310. The browser interprets the HTML file and builds the web page with source (i.e., from the origin web server) or cached images. The cached images can be available locally or at an ISP, or at a router or other network device along the path. If in decision block 304, it is determined that the page is not cached at the client, then another test is performed in decision block 306 to determine if the page has been cached at an ISP. The ISP cache test is intended to be illustrative of an embodiment of the invention. The next hop from the client can be to a server on an intranet which has a TCP/IP address and provides direct Internet access. If the page has been cached along the path, then, as indicated in logic block 312, the HTML file is delivered to the client browser to interpret the HTML code and build the web page with images that have ben cached or retrieved from the origin web server. If the page has not been cached along the path to the web server, the request for the page is transmitted to the host where the web server software processes the request as indicated in logic block 308. If the browser has requested an HTML file, the web server retrieves the original

source HTML file, attaches a header to the file, and send the file to the browser as indicated in logic block 314.

Sub
AI
5 Once the browser has received the HTML file from the processing in logic blocks 310, 312 or 314, a test is made in decision block 318 to determine if the HTML file contains an uncacheable single pixel GIF (represented by uc.GIF in this invention). If it does not, the retrieved cached images are displayed to complete the build of the web requested web page. Processing of the request is then completed as indicated by termination block 326. If, in decision block 318, a uc.GIF request is found in the HTML file, then the uc.GIF and CGI query string are transmitted to the origin web server where they are analyzed to gather the enriched web server activity data made possible by this invention. The browser again interprets the HTML code and builds the page with source or cached images. Using the example of Fig. 2, 14 hits are recorded for the web page, including one for the transmitted uc.GIF request and 13 for the other source images that are retrieved based on the HTML IMG SRC tags/attributes in the HTML file. This represents the surrogate nature of using the uncacheable single pixel GIF request. The referent page for the 14 hits is also contained as part of the CGI query string. In Fig. 2, this is represented by "r = '&ref='+top.document.referrer". The gathering and storing of this enriched web server activity data is indicated by logic block 322. The request processing then ends as indicated in termination block 324.

20 When a user visits a website, the browser examines the URL and looks into a cookie file stored on the client computer's hard drive. If the browser finds a cookie associated with that URL, it sends that cookie information to the server. If no cookie is associated with the URL, the server places a cookie inside the cookie file. Some sites may first ask a series of questions, such as name and password, and then will place a cookie on the hard disk with that information in it. This is

typical of sites that require registration. Commonly, a GCI script on the server takes the information that the user has entered and then writes a cookie onto the client computer's hard disk. When the user leaves a web site, the cookie information remains on the hard disk so that the site can recognize the user the next time the user visits the web site, unless the cookie has specifically been written to expire when the user leaves the site.

With the capability to gather enriched information through the use of the single pixel GIF described above, much more detailed and accurate information regarding web site activity can be collected and stored in multidimensional databases, including multidimensional implementations of a relational database. Furthermore, this collected data also can be analyzed using relatively new techniques such as On-line Analytical Processing (OLAP), described briefly below.

On-Line Analytical Processing (OLAP) describes a class of technologies that are designed for live ad hoc data access and analysis. While transaction processing generally relies on relational databases, OLAP has become synonymous with multidimensional views of business data. These multidimensional views are supported by multidimensional database technology. OLAP applications are used by analysts who frequently want a higher level, aggregated view of the data, such as total sales by product line, by region, etc. The OLAP database is usually updated in batch mode, often from multiple sources, and provides an analytical backend to multiple user applications.

Fig. 4 illustrates an exemplary site level analysis display that can be derived from the collecting of accurate hit information using the single pixel GIF as a surrogate for the complete set of log records which would have been generated if the web page content had not been cached. The figure depicts the various measurements that can be made for selected intervals of time and includes

hits, pages visited, seconds per page view, visits, hits per visit, page views per visit, and seconds per visit.

Fig. 5 illustrates an exemplary referral categories display that can be generated from the use of the single pixel GIF to log information pertaining to the web page referral source. The different referral categories include commercial, education, government, internal referrals, ISP referrals, and search engines and directories among them. Again, the data is presented for selected intervals of time (e.g., calendar weeks). The various referral categories are underlined, which means that they can "drilled down" to sub-referral categories as illustrated in Fig. 6.

Fig.6 illustrates the breakdown of the search engines and directories referral category for the selected intervals of time based on the referrals made from common search engines or browsers. For example, during the week ending June 10 in which the peak number of page referrals occurred, over 71% were referred by the Yahoo search engine. Further drill down is possible into the search engine referral category as indicated by the underlined subcategories.

Fig. 7 illustrates a further drill down of the AltaVista referral subcategory. For example, the display shows that 84% of the referrals from AltaVista during the week ending June 3 originated from a CGI query string on the AltaVista home page. No further drill downs are possible in this referral subcategory.

Fig. 8 illustrates an exemplary display of web page by content categories that can be derived from the collecting of accurate hit information using the single pixel GIF as a surrogate for the complete set of log records which would have been generated if the web page content had not been cached. The content categories include draws, home page, news and photos, players, scoreboard, and shop (gift shop) among other content categories. The data is presented for selected intervals of time.

The various content categories are underlined which means they can be drilled down to a lower level of detail.

Fig. 9 illustrates a drill down of the home page content category. The resources include the English version home page (/en) accessible via a Java Script-enabled browser; the French version home page (/fr) accessible via a Java Script-enabled browser; the English version home page (/en/index.html) accessible from a browser that is not Java Script-enabled, etc. For the peak traffic week ending June 10, 58% of the home page traffic was directed to the English-version page and initiated from a Java Script-enabled browser. Slightly less than 42% of the traffic was directed to the French-version page initiated from a Java Script-enabled browser.

Fig. 10 illustrates a display of exemplary saved reports that can be generated using OLAP processing of the surrogate log records created through the use of the single pixel GIF of this invention. The saved reports include site level reports, visit distribution reports, traffic reports, content reports, domain/sub-domain reports etc. Each of the listed reports is underlined indicating that a detailed report is available simply by clicking on the report name.

Figs. 11A-11M illustrate the format of the corresponding exemplary saved report. Fig. 11A shows the site level report that is available. In this instance, the available site level report is a site traffic report. The report name is underline indicating that a further drill down to a detailed report results from clicking on the report name. Such action would generate a display like that of Fig. 4. The available visit distribution reports are listed in the display of Fig. 11B. Figs 11C-11K and 11M illustrate various saved reports that are basically "top 10" lists. Fig. 11C depicts traffic reports and enables display of the top 10 requested resources. Fig. 11D depicts content reports and enables display of the top 10 most requested pages. Fig. 11E depicts sub-domain reports and enables

display of the top 10 sub-domains by either pages viewed or by number of visits. Fig. 11F depicts domain reports and enables display of the top 10 domains by either pages viewed or by number of visits. Fig. 11G depicts referral reports and enables display of the top 10 referrals by pages viewed or by number of visits. Fig. 11H depicts entry page reports and enables display of the top 10 site entry pages. Fig. 11I depicts exit page reports and enables display of the top 10 exit pages. Fig. 11J depicts browser reports and enables display of the top 10 browsers by either pages viewed or by number of visits. Fig. 11K depicts platform reports and enables display of the top 10 platforms by pages viewed or by the number of visits. Fig. 11L depicts usage cluster reports and enables display of usage cluster visits. Fig. 11M depicts ad reports and enables display of the top 10 ads by impression created. All of the available saved reports are presented for selected intervals of time such as the most recent five weeks.

The corresponding structures, materials, acts, and equivalents of any means plus function elements in any claims below are intended to include any structure, material, or acts for performing the functions in combination with other claimed elements as specifically claimed.

While the invention has been particularly shown and described with reference to embodiments thereof, it will be understood by those skilled in the art that various changes in form and detail may be made without departing from the spirit and scope of the present invention.